華語文教學研究 10.2:117-141 , 2013

The Perception of Second Language Fluency and Foreign Accents*

Chen-huei WU
National Hsinchu University of Education
Department of Chinese Language and Literature

Abstract

This study examined the factors contributing to perceptions of second language fluency and foreign accents by randomly selecting speech samples from the Spontaneous Chinese Learner Speech Corpus for perceptual ratings. Eight rating questions that reflected the oral proficiency of fluency, nativeness, accentedness, disfluencies, pronunciation, grammar, vocabulary, and comprehensibility were evaluated. The correlation analysis revealed that the fluency rating was correlated with disfluencies and vocabulary and accentedness correlated with nativeness and pronunciation. The principal component analysis showed that all the rating variables contributed similar weights to the first principal component (PC1), which represents the general impression of oral proficiency. The PC2 classified the rating variables into two categories. One group consisted of the knowledge factors of fluency, disfluency, grammar, vocabulary, and comprehensibility, whereas the other group included the sound-related factors of nativeness, accentedness, and pronunciation. The PC2 was mainly composed of nativeness and accentedness. The PC3 stood for the

* This work was enabled by the Critical Research Initiatives Grant, University of Illinois at Urbana-Champaign, and the NSF project "DHB: An Interdisciplinary Study of the Dynamics of Second Language Fluency." The rating survey was conducted with the 2010 Jiede Empirical Research Grant, CLTA. We thank the PIs, Chilin Shih, Mark Hasegawa-Johnson, J. Kathryn Bock, and Fred Davidson and the raters, students, and Chinese TAs in the Chinese classes for participating in this project. We thank the anonymous reviewers for their comments and suggestions. Any opinions, findings, and conclusions expressed in this paper are those of the author.

characteristics of oral fluency. Further multiple regression analysis showed that disfluencies and vocabulary have a substantial causal effect on fluency, whereas nativeness has a substantial effect on accent. The results have implications for the relationship between speech planning and fluency based on the Levelt speech production model.

Keywords: second language fluency, foreign accent, speech perception

1. Introduction

This study addressed second language fluency, foreign accents, and whether it is possible for an adolescent or adult learner of a second language (L2) to speak with an accent in a fluent manner. L2 learners have varying degrees of fluency and accents; light accents do not considerably affect communication. Heavy accents may lead to inefficient communication or miscommunication, which may result in unemployment (Hosoda & Stone-Romero, 2010), less credibility and persuasiveness (Lev-Ari & Keysar, 2010), and stereotypes and bias (Lindemann, 2005). This occurs because understanding a person with disfluent or accented speech is difficult and requires more processing time for listeners to comprehend. Therefore, understanding the characteristics of non-native speech is critical to designing a corrective method to help language learners to attain higher fluency levels and accent reduction.

This study examined L2 fluency and foreign accents in spontaneous speech produced by Chinese instructors, heritage speakers,[1] and English-speaking learners of Chinese. Speech data were selected from a large corpus; that is, the Spontaneous Chinese Learner Speech Corpus (Shih, 2006), which covers varying degrees of fluency and foreign accents. The major research question is as follows: What leads to the perception of L2 fluency and foreign accents? This question was addressed by collecting perceptual ratings from 43 untrained native Manda-

[1] Heritage speakers are students whose native language is Chinese, but who have received education in English and were raised in the United States. Many were born in the United States, whereas others arrived in the United States at a young age.

rin listeners in Taiwan. The rating questions included aspects of fluency, nativeness, accentedness, disfluency, pronunciation, grammar, vocabulary, and comprehensibility. The ratings were subjected to correlation analysis, principal component analysis (PCA), and multiple regression analysis to examine the relationship between the rating variables and L2 fluency and foreign accents.

The remainder of this paper is organized as follows: Section 2 presents the relationship between fluency and the speech formation model and previous studies related to the perception of foreign accents; Section 3 introduces the methodology, including the corpus, sampling design, and rating experiment; Section 4 details the results and provides the discussion; and finally, Section 5 offers a conclusion.

2. Literature Review

2.1 Fluency and Speech Planning

Lennon (1990; 2000) defined fluency in two senses: a broad sense, corresponding to all-around oral proficiency; and a narrow sense, referring to the speech rate and smoothness indicated by disfluency, such as filled pauses, hesitation, repetition, and lengthened syllables. In practice, features defined by a narrow sense of fluency are measurable and can be detected by listeners. The rating experiment focused on perceived fluency, which uses the perspectives and the impressions of the listeners of the speakers. We also examined factors that contribute to perceived fluency.

The Levelt model (Levelt 1989) has a considerable influence on contemporary ideas of speech production, and is a model of the mental process required for the creation of fluent speech in first language (L1) speech production. Three main modules comprise production: the conceptualizer, formulator, and articulator modules. In the conceptualizer module, the content of the intended message is developed as a form of a preverbal message by speakers. The output of the conceptualizer is subsequently sent to the formulator to extract the appropriate meaning and forms of lexical items to express the message. Grammatical encoding forms the surface syntactic structure and relays the message for phonological encoding. A phonetic plan is delivered to the articulator to produce overt speech.

Thus, fluency reflects multiple rapid and efficient processes in speech planning, including lexical access, grammatical and phonological encoding, and the formation of speech into articulatory output.

De Bot (1992) argued that multilingual speakers have only one mental lexicon, whereas each language has its own module linked to lexical entries. This implies that more than one linguistic system competes with other systems and lexical retrieval may go through L1 module or cross other linguistic modules. This may slow down the formation of speech. Lennon (2000) suggested that linguistic proficiency involves linguistic knowledge (phonology, vocabulary, and grammar), pragmatic knowledge (form-function matching), and processing skills. Deficient fluency mainly results from processing or memory-storage demands rather than knowledge. However, whether disfluency results from a lack of knowledge or from the processing skills is difficult to distinguish from speaker performance. Although perceived fluency may not reflect speaker knowledge or processing, listeners can detect filled pauses in speech that the speaker may be unaware of.

2.2 Perceptions of Foreign Accents

Various studies have indicated that several objective variables are related to the perception of accentedness, including segmental and prosodic errors, speaking rate, and the background of the listener (Anderson-Hsieh et al. 1992; Munro 1998; Munro & Derwing 2001; Munro et al. 2006; Lindemann 2005). In addition, numerous studies have examined the relationship among perceived fluency, foreign accents, and suprasegmental properties in L2 speech. Anderson-Hsieh et al. (1992) indicated that inaccurate L2 production of prosodic properties, such as stress, rhythm, and intonation, may contribute to the pronunciation ratings more strongly than inaccurate L2 segmental production. Trofovimich and Baker (2006) examined several suprasegmental features (stress timing, peak alignment, speech rate, pause frequency, and pause duration) in the production of adult Korean L2 learners of English and the manner in which each feature contributed to fluency and foreign accents. The results showed that the amount of L2 experience influenced the production of stress timing, whereas the onset age of intensive L2 ex-

posure influenced the other factors (speech rate, pause frequency, and pause duration). Moreover, only pause duration and speech rate have considerable predictive power on the perception of foreign accents.

L2 learners tend to speak slower than native speakers do. This cannot be regarded as a transfer effect from L1 because not all L1 learners have a quicker speaking rate than L2 learners do. Several factors cause L2 learners to speak at slower rates, including articulatory difficulties in segmental and prosodic accuracy, slower lexical retrieval, and incomplete syntactic and morphological knowledge (Munro & Derwing 2001: 453). Flege (1988) examined accentedness ratings after removing pauses from sentences that were read by Mandarin and Taiwanese learners of English, and did not observe significant differences in accent ratings between normal and modified conditions because the majority of the removed pauses were short in duration (approximately 200 ms). Munro and Derwing (1998) manipulated the speech rates of the speech production of 10 Mandarin learners of English to examine their effects on their listeners' judgments of accentedness. The original rate was compressed to a 10% faster rate and also stretched to a 10% slower rate. They observed a curvilinear relationship between speaking rates and accentedness ratings; that is, fast and slow speaking rates increased accentedness and decreased comprehensibility. Munro and Derwing (1998) indicated that fast speech required extra processing and slow speech resulted in retention of information in short-term memory for a longer period. Both types of speech caused processing difficulties and resulted in lower accent ratings. They found that the optimal speaking rate for L2 speech was slightly faster than that of L2 speakers but slower than the typical speaking rates of native speakers. Although the definition of optimal speaking rates for L2 learners is unclear, this result suggested that the speaking rate influences foreign accents in addition to L1 transfer.

3. Methodology

3.1 Spontaneous Chinese Learner Speech Corpus

A large corpus, the Spontaneous Chinese Learner Speech Corpus, was used to examine perceptions of L2 fluency and foreign accents. The corpus consists of

185 hr of audio and video recordings from third-year and fourth-year Chinese language classes. The recordings were conducted in a Chinese speech-training class on a weekly basis from autumn 2004 to spring 2009 at the University of Illinois at Urbana-Champaign (Shih 2006; Shih & Wu 2011; Wu 2011). The speakers in the corpus included 11 Chinese teachers (9 women and 2 men), 86 Chinese-heritage learners (28 women and 58 men), 11 Korean learners of Chinese (7 women and 4 men), and 23 English-speaking learners of Chinese (8 women and 15 men). The heritage learners were students whose native language is Chinese, but who received education in English and were raised in the United States. Many were born in the United States, whereas others arrived in the United States at a young age. Most Chinese-heritage learners were from mainland China, but some were from Taiwan. The non-native L2 learners were students whose native language is English or Korean. The English-speaking learners of Chinese were learners who had no prior background in Chinese before attending college-level Chinese classes. In contrast to the English-speaking learners of Chinese, most of the Korean learners had a prior background in Chinese during their high-school education.

Students in the Chinese classes received speech training in two paradigms: a variety show and a debate (Shih 2006). Each paradigm was designed for a 50-min class. The variety-show format comprises four main sessions: opening, talk show, formal speech, and comments. Learners are asked to play roles, such as the chair for the entire show, the talk show host, or speech-makers. The chair opens each session in the show with an introduction, and the talk show host prepares several topics and selects students from the audience to stand in front of the stage and answer questions. The speech-makers deliver prepared formal speeches of 4-6 min in duration. The variety-show format incorporates a few frequently encountered social interactions, such as delivering opening or closing remarks, introducing guest speakers, and delivering formal speeches. Through weekly practice, learners have multiple opportunities to play each role and observe the performances of their classmates and instructors.

In the debate format, students are divided into two sides: a proposition side and an opposition side. A specific topic is provided in advance. Certain learners

prepare formal speeches to express their positions on the specific topic, some prepare questions to ask the opposing side, and some must answer questions on the spot. The debate format trains learners to argue and speak clearly, logically, and convincingly under time pressure.

Based on various formats, two speech styles are used: (a) spontaneous speech, in which students speak without advanced preparation, such as certain questions and all answers in both the variety shows and the debates; and (b) prepared speech, such as speeches delivered by the chair or host, formal speeches prepared by students for the variety shows, and statements made by students in the debates. If students prepare their speeches in advance, most would read their speeches while holding their drafts. Thus, prepared speech can be recognized in video clips.

Speaker periods were marked and annotated by trained research assistants using ELAN video editing software (Hellwig, n.d.) Each speaker period serves as a unit for sampling individual speakers. Overlapped speech by multiple speakers was eliminated because it increases difficulty in perceptual ratings. Speaker periods were marked from the beginning of the period to preserve discourse coherence. The length of speaker periods varied from a few seconds to less than 3 minutes.[2]

3.2 Sampling Design

A subset of the data was selected from the corpus to obtain perceptual judgments of the fluency of foreign accents. Fluency was determined by the manner in which speakers managed connected sentences. Speech samples must be sufficiently long to allow raters to evaluate fluency. In addition, each speech sample must include multiple sentences instead of sentence fragments or single sentences, if possible. However, the length of speech samples is not universally agreed-on for perceptual ratings. Ambady and Rosenthal (1993) showed that end-of-semester teaching evaluations were successfully predicted by the stu-

[2] Long speech files were cut into short snippets for other research purposes, such as automatic speech recognition. Long files increase the chance of misalignment between speech and text.

dents' ratings of the nonverbal behaviors of the instructor based on 30 s of silent video clips composed of three 10 s clips from the same teacher or smaller clips of 6 s and 15 s. This result suggested that impressions can be formed quickly. Derwing et al. (2006) used 20-s speech samples for evaluating fluency and foreign accents and observed that 20 s was sufficient for raters to make reliable judgments. However, trade-offs between the length of the speech samples and the duration of the experiment are inevitable. Using longer speech samples increases the duration of the experiment and the demands required of the raters. Another limitation of longer speech samples is the difficulty of obtaining long spontaneous speech from language learners if their oral proficiency is low.

Because of these concerns, 1 minute of speech for each speaker was composed of four 15-s snippets at various times in a spontaneous speech style (mainly from the questions and answers in the variety show) that was randomly selected from the corpus.

For native speakers, 11 Chinese instructors (9 women and 2 men) who fully acquired their L1 in Mandarin served as the baseline for comparing the results with heritage and English-speaking learners of Chinese. Four 15-s snippets were randomly selected from the database. For language learners, 15 recordings of class sessions were conducted for each semester and divided into three blocks. Two snippets were chosen from the first block at the beginning of the semester (the first 5 weeks) and two snippets were chosen from the last block at the end of the semester (the last 5 weeks). A 4- or 5-week gap occurred between the blocks at the beginning and the end of the semester. If a learner attended classes for more than one semester, 1 min was chosen from each semester. The speech samples of 17 heritage speakers (5 women and 12 men) and 20 English-speaking learners of Chinese (5 women and 15 men) were randomly chosen based on the block design. A total of 236 speech files were chosen for ratings and analysis.

3.3 Perceptual Rating Design

Forty-three native speakers of Mandarin in Taiwan rated the 236 snippets. All the raters were untrained and linguistically naive undergraduate students at

the National Changhua University of Education and National Chiao-Tung University.

The rating was conducted through a web interface. The rating snippets were divided into 6 sessions and were presented pseudo-randomly in each session. Eight questions were presented in two pages; the audio files were played at least once for each page. One training snippet was provided at the beginning of each session. The sound files were auto-played when raters entered the question pages and could be played as many times as required. After raters submitted the answer and entered the next page, they were not allowed to return to the previous page to change their answers.

The eight questions listed in (1) represent speech performance regarding fluency, nativeness, accentedness, disfluencies, pronunciation, grammar, vocabulary, and comprehensibility.

(1) Rating questions

a.Page 1

Fluency: Was the speaker fluent?

Nativeness: Did the speaker sound like a native Chinese speaker?

Accentedness: How accented was the speech?

b.Page 2

Disfluencies: Did you notice many pauses, repetitions, or hesitations?

Pronunciation: Was the speaker's pronunciation understood?

Grammar: Did the speaker use sentence structures incorrectly?

Vocabulary: Did the speaker have difficulty in selecting appropriate vocabulary to express him/herself?

Comprehensibility: Could the speaker convey his/her intended message?

Each snippet was rated on a binary scale, 1 (*not fluent*) or 2 (*very fluent*), for the fluency rating and a 4-point scale for the remaining criteria. Higher scores indicated more positive values of the variables. For example, higher scores of nativeness indicate that the speech is more native-like. The higher scores of accentedness indicated lighter accents of the speakers. The design of binary fluency

ratings can be used for future development of automatic assessment systems of fluency. The objective is to set a threshold to classify speakers into 2 groups: fluent and not fluent. A yes/no question must be used for nativeness based on speaker identities. However, determining whether heritage learners are native speakers is difficult. Moreover, examining the manner in which listeners perceive a speaker as native, non-native, or somewhere between these two categories is essential. Accentedness is a rating to measure the perceptual distance of speech between speakers and listeners, such as dialect accent and foreign accent. Dialect accent refers to language differences in various regions. For example, the speech of people living in Hong Kong or mainland China differs from that of people living in Taiwan regarding pronunciation, vocabulary, and grammar. When people speak a dialect in a particular region, they are perceived as having an accent by speakers from other regions. Similarly, when people learn a second or third language, their speech may differ from that of native speakers. This is called a foreign accent. Pronunciation, grammar, and vocabulary are usually used as criteria in language testing or for language instructors to evaluate the performance of learners in classroom settings. The grammar rating is used to measure the correctness of sentence structure. The vocabulary rating indicates the vocabulary size of the speakers and whether they can select appropriate words. Disfluency is a quantifier that is widely used to measure pauses, silence, self-corrections, repairs, and repetition in speech. This rating is used to gain a sense of disfluencies in perception. Comprehensibility is used to determine whether a rater can easily comprehend the messages delivered by speakers.

4. Analyses and Discussion

A two-way repeated measure ANOVA was conducted to determine whether any improvement of the rating scores occurred between the beginning and end of the semester using the semester period (2 levels) and 8 rating variables (8 levels) as within-subjects factors. No significant difference was observed in the English-speaking learner group ($F = 1.783$, $p = .193$), whereas the heritage group exhibited a slight difference ($F = 5.84$, $p = 0.025 < 0.05$). Further analysis indicated that the effect size based on the mean rating scores of the heritage group

was small (approximately 0.1 – 0.3). In addition, learning was not obvious because the time between the beginning and end of the semester was insufficient to observe differences. Thus, data of various semester blocks were combined for further analysis.

4.1 Correlation Analysis

Tables 1, 2, and 3 show the correlation matrices of rating variables by Mandarin native speakers, heritage learners, and English-speaking learners, respectively. The rating variables are highly correlated in all 3 groups, and all correlations were significant at .001. This may be attributed to the possibility that all variables correlated with proficiency in Chinese. For example, accentedness and grammar correlated with proficiency; thus, they correlated with each other. In other words, all variables may not be directly related; however, they may be excellent predictors of fluency and foreign accent because of their potential correlation with proficiency.

The correlations were stronger in the heritage group than those of the native and English-learner groups. To determine whether the correlations differed significantly among speaker groups, a Fisher z' transformation of the correlation was performed and the difference was computed between samples of various sizes. The results showed that the correlations of the heritage group did not differ significantly from those of the native-speaker group; however, they did differ from those of the English-learner group. The correlations of the English-learner group differed significantly from those of the native group.

Table 1: Correlation matrix of the rating variables of Mandarin native speakers. All correlations are significant at the .001 level.

corr (r)	Fluency	Native	Accent	Disfl.	Pron.	Grammar	Vocab.	Comp.
Fluency	1	0.86	0.54	0.72	0.84	0.79	0.87	0.84
Native		1	0.67	0.77	0.88	0.88	0.89	0.86
Accent			1	0.45	0.72	0.63	0.59	0.66
Disflu.				1	0.71	0.68	0.87	0.77
Pron.					1	0.92	0.88	0.94
Grammar						1	0.86	0.90
Vocab.							1	0.87
Comp.								1

Table 2: Correlation matrix of the rating variables of heritage learners. All correlations are significant at the .001 level.

corr (r)	Fluency	Native	Accent	Disfl.	Pron.	Grammar	Vocab.	Comp.
Fluency	1	0.88	0.79	0.95	0.87	0.91	0.95	0.91
Native		1	0.94	0.85	0.91	0.87	0.88	0.86
Accent			1	0.79	0.91	0.83	0.83	0.83
Disflu.				1	0.86	0.91	0.96	0.92
Pron.					1	0.93	0.92	0.95
Grammar						1	0.96	0.97
Vocab.							1	0.97
Comp.								1

Table 3: Correlation matrix of the rating variables of English-speaking learners. All correlations are significant at the .001 level.

corr (r)	Fluency	Native	Accent	Disfl.	Pron.	Grammar	Vocab.	Comp.
Fluency	1	0.62	0.45	0.95	0.66	0.82	0.93	0.83
Native		1	0.93	0.56	0.80	0.69	0.68	0.68
Accent			1	0.37	0.79	0.58	0.51	0.58
Disflu.				1	0.59	0.82	0.96	0.84
Pron.					1	0.83	0.71	0.85
Grammar						1	0.90	0.95
Vocab.							1	0.92
Comp.								1

The results of the correlation analysis are summarized, as follows: first, the ratings of fluency, disfluency, and vocabulary were highly correlated. Second, the nativeness rating was highly correlated with the accentedness rating of the heritage and English-learner groups. For native speakers, the nativeness rating correlated with the ratings of pronunciation, vocabulary, and comprehensibility. Third, the accentedness rating was highly correlated with the pronunciation rating. Fourth, the ratings of pronunciation, grammar, vocabulary, and comprehensibility were highly correlated with each other.

The correlation analysis showed that all rating variables were highly correlated with each other; however, the magnitude of the correlations differed. Thus, the pattern of correlation analysis among speaker groups was examined. The following figures show the rating scores as mean values of each snippet averaged from 43 raters. Figure 1 shows the correlation between fluency and accentedness among the speaker groups (native: $r = 0.54$; heritage: $r = 0.79$; English: $r = 0.45$, $p < .001$). The magnitude of the linear fit is low. The heritage-learner group exhibited the strongest relationship. Numerous learners received high-fluency ratings (e.g., 1.65) with low-accent ratings (lower than 2). Other learners received low-accent ratings (approximately 2.5) with relatively low-fluency ratings (e.g., 1.3). The trend of the data distribution was not linear, indicating that a language learner can have a strong accent and be judged as fluent, whereas other learners

have mild accents and are not perceived as fluent.

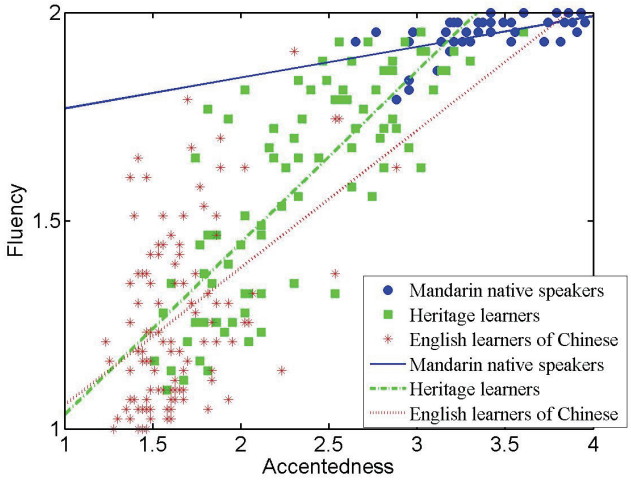

Figure 1: Correlation between fluency and accentedness

Figure 2 shows the correlation between accentedness and pronunciation among speaker groups (native: r = 0.88; heritage: r = 0.91; English: r = 0.88, $p < .001$). The data distribution indicates that English-speaking learners can gain high-pronunciation ratings (2.5 to 3), whereas their accentedness ratings remain low (1.5 to 2). Similarly, heritage learners can achieve pronunciation ratings between 3.5 and 4 and accentedness scores between 3 and 3.5. This suggests that improving pronunciation is easier than improving people's impressions of an accent.

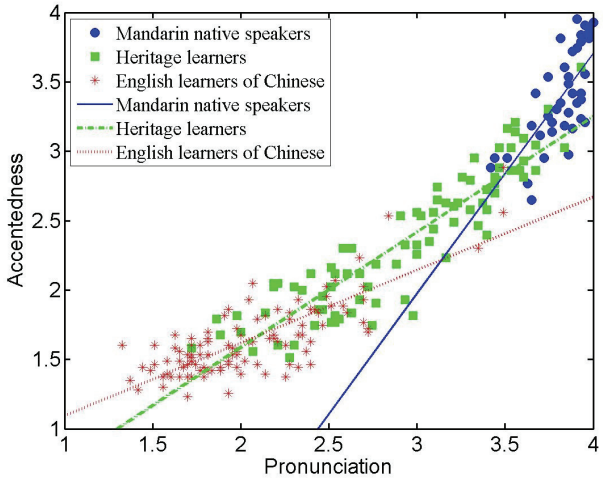

Figure 2: Correlation between Accentedness and Pronunciation.

Figure 3 shows the correlation between nativeness and accentedness of the classroom data among speaker groups. Accentedness correlated strongly with nativeness, especially in the learner group (native: r = 0.67; heritage: r = 0.94; English: r = 0.93, $p < .001$), indicating that higher nativeness scores result in lower perceptions of accent. Native speakers were perceived as native with scores ranging from 3.4 to 4, whereas their accent scores were between 3 and 4.

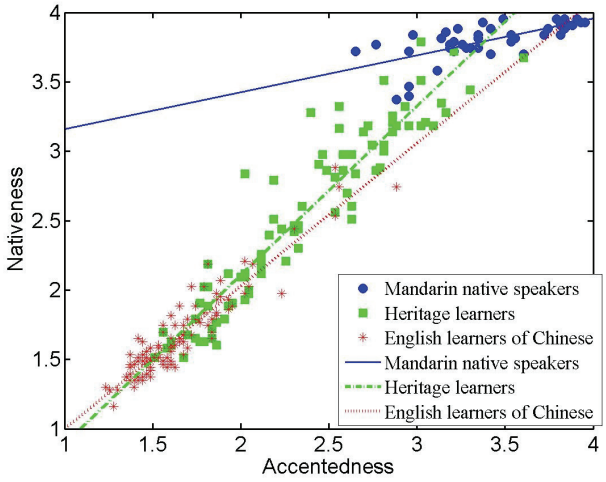

Figure 3: Correlation between Nativeness and Accentedness.

Table 4: Mean rating scores of fluency, nativeness and accentedness in the native group.

Speakers	Gender	Originality	Fluency	Native	Accent
LHY	F	Taiwan	1.97	3.90	3.81
SJL	F	Taiwan	1.97	3.85	3.39
STJ	F	Taiwan	1.98	3.88	3.70
WZH	F	Taiwan	1.96	3.90	3.88
ZYX	M	Taiwan	1.98	3.86	3.80
GJ	F	Mainland China	1.89	3.66	3.10
GY	F	Mainland China	1.97	3.85	3.17
JY	F	Mainland China	1.95	3.77	3.09
WYJ	F	Mainland China	1.95	3.77	3.28
WS	F	Mainland China	1.95	3.77	3.44
LTL	M	Mainland China	1.87	3.59	2.94

The mean scores of fluency, nativeness, and accentedness were verified to further examine the ratings of the native group, as shown in Table 4. The table shows that native speakers from Taiwan received higher accent scores (ranging from 3.4 to 3.88) than speakers from mainland China (ranging from 2.9 to 3.4), whereas the fluency and nativeness ratings were high in all native productions.

This occurred because the listeners were from Taiwan and detected the differing dialectal accents of speech from mainland China and ranked it lower; however, they provided higher rankings to speech that was closer to their own speech. The Taiwanese speaker (SJL) with the lowest accent ranking was in the United States longer than the native Chinese speakers were, which may be a factor leading to differences in speech. This result provides evidence of the conceptions of accents, which are defined as the perceptual distance of speech between listeners and speakers.

4.2 Principal Component Analysis

The rating variables were not independent because they were highly correlated with each other. This suggests that more than one variable may be measuring the same behavior. Therefore, PCA was used for reducing dimensionality and

determining the internal structure of the data. PCA is a type of exploratory factor analysis (EFA), which is a statistical model used to explore a reduced number of unobserved variables in the absence of an assumption or hypotheses regarding the construct of the measures. EFA determines the number of factors and selects the extraction and rotation methods. In PCA, the variance is maximized and all variability in the data is considered. When all rating scores are plotted together in a multi-dimensional space, the PCA model rotates the entire data set and establishes the optimal viewing angle to visualize the data set. Rather than using the original 8 rating variables to predict the quality of speech, the PCA model generated a new set of components based on the coefficients of the original variables. The PCA model creates the same number of principal components as the original variables (e.g., the PCA model creates 8 principal components based on the 8 rating variables); however, the first two components are usually sufficient to explain the variance in the data. Each component is a linear combination of all original variables with differing weights; the components were formulated independently during analysis.

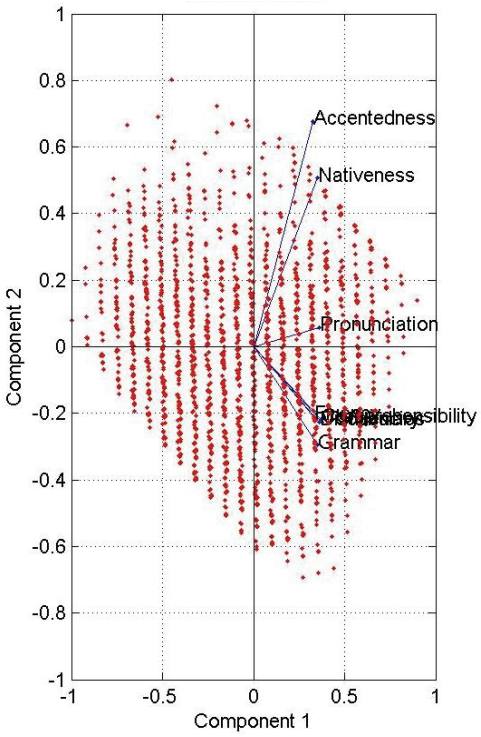

Figure 4. PCA of rating scores

As shown in Figure 4, the first principal component (PC1) is represented by the horizontal axis and the second principal component (PC2) is represented by the vertical axis. The rating experiment yielded 81,184 rating scores (8 questions x 43 raters x 236 snippets) of the data. Each variable is represented by a vector. The length and direction of each vector indicate the manner in which each variable contributes to the two PCs; that is, the projection of vectors on the axes is the weight of the original variables in the linear combination of the PCs. Vectors pointing in the same direction indicate variables that are correlated. Long variable lines indicate a strong correlation between an axis and the corresponding variable.

All rating variables contributed to the PC1 with similar weights, confirming the results that were obtained from the correlation matrix. The equation of the PC1 is formulated in (2a), which is the linear combination of all variables. Each variable contributed to the weight (coefficients) ranging from 0.33 to 0.37. The weight is the projection of each variable vector on the x-axis (PC1). The PC1

represents the general impression of oral proficiency.

(2) Equations of PC1, PC2 and PC3:

a. PC1 = 0.33*Fluency + 0.35*Native + 0.33*Accent + 0.36*Disflu + 0.36*Pron + 0.35*Grammar + 0.37*Vocab + 0.37*Comp

b. PC2 = -0.21*Fluency + 0.5*Native + 0.68*Accent - 0.23*Disflu + 0.06*Pron - 0.3*Grammar - 0.23*Vocab - 0.22*Comp

c. PC3 = 0.76*Fluency + 0.08*Native + 0.04*Accent + 0.27*Disflu - 0.39*Pron - 0.35*Grammar - 0.07*Vocab - 0.27*Comp

The PC2 is a dimension containing the sound-related factors of accentedness, nativeness, and pronunciation, and the knowledge-related factors of fluency, disfluency, grammar, vocabulary, and comprehensibility. The sound-related factors (positive coefficients, weight, and projection on the y-axis) are a measure of the quality of pronunciation, whereas the knowledge factors (negative coefficients, weight, and projection on the y-axis) are a measure of whether speakers can form speech and express ideas clearly. This indicates that linguistic knowledge, which resembles the building blocks of speech, is related to L2 fluency (as addressed in the Levelt model) and may affect perceived fluency. The lines for the sound-related factors and knowledge-related factors are nearly perpendicular, indicating that they do not behave similarly. The equation of the PC2 is shown in (2b). As we can see, the variables nativeness (0.5) and accentedness (0.68) contributed the major weights to the PC2. Thus, the PC2 stands for the color of nativeness and accentedness in oral performance.

In (2c), the variable fluency contributed the primary weight 0.76 to the PC3. The weights contributed by other variables in the PC3 were either small or negative. Therefore, the PC3 represents the characteristics of fluency in oral speech.

Figure 5 shows that the break of the amount of variance explained by each component occurred between the PC1 and PC2. The PC1 explains 72.7% and the PC2 explains 8% of the variance in the data. Combining PC1 and PC2 can explain approximately 80% of the variance. The PC3 can explain 5.8% of the variance, which yields significance at 0.5.

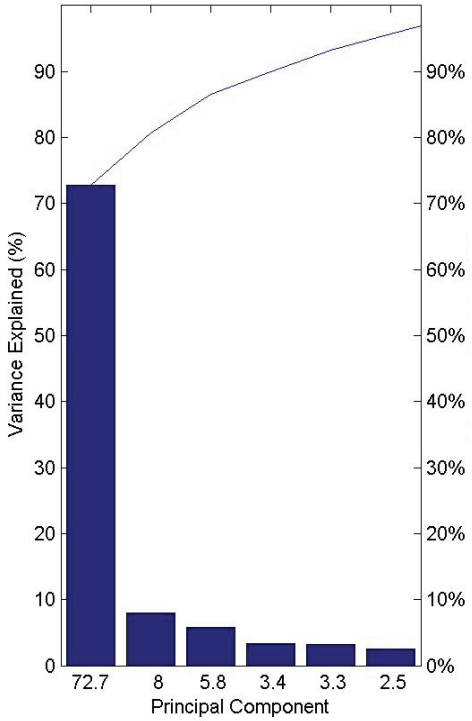

Figure 5. The percent of total variability explained by each component.

4.3 Multiple Regression Analysis

Based on the results of grouping the variables in PCA, the data were analyzed using the multiple regression method to examine the causal effects of the independent variables on the dependent variables (fluency and accent). The knowledge-related factors of disfluencies, grammar, vocabulary, and comprehensibility were treated as independent variables to predict the dependent variable (fluency).

The ANOVA F-value was significant ($F = 810.602$, df(regression) = 4, df(residual) = 43, $p < .001$), indicating that the regression model was an excellent fit. Among the 4 variables, the coefficients of disfluencies and vocabulary were significant (disfluencies: $t = 2.06$, $p < .05$; vocabulary: $t = 3.15$, $p < .01$). When the other variables were constant, fluency scores were positively related to disfluencies and vocabulary, increasing by 0.11 for less disfluencies and 0.31 for extra vocabulary. Finally, 98.7% of the variance in fluency was explained by dis-

fluencies, grammar, vocabulary, and comprehensibility ($R^2 = .987$).

The sound-related factors of nativeness and pronunciation were used to predict accent. The ANOVA F-value was significant (F = 530.591, df(regression) = 2, df(residual) = 45, $p < .001$), indicating that the regression model was an excellent fit. Among the 2 variables, the coefficients of nativeness were significant ($t = 5.06$, $p < .001$). With a constant pronunciation variable, the accent scores were positively related to nativeness, increasing by 0.65 for every higher score of nativeness. Consequently, 95.9% of the variance in accentedness was explained by nativeness and pronunciation ($R^2 = .959$).

The results of the multiple regression analysis indicated that disfluencies and vocabulary have significant predictive power on fluency and nativeness has a significant causal effect on accent.

5. Conclusion

This study examined the factors contributing to perceived fluency and foreign accents by conducting a rating experiment on randomly selected samples from a speech corpus. Forty-three linguistic native raters listened to each speech sample and answered eight questions on their impression of the speech. They first reported their broad impression regarding whether the speech sounded fluent, whether the speaker had an accent, and whether the speaker was a native Chinese speaker. Subsequently, they evaluated specific aspects of the speech sample, including perceived disfluencies, efficiency of pronunciation, correctness of grammar, appropriateness of vocabulary selection, and ease of comprehension. A correlation analysis, PCA, and multiple regression analysis were conducted to determine the factors that influence the perceptions of native listeners of L2 fluency and foreign accents.

The correlation analysis showed that all rating variables were correlated, indicating that fluency is highly correlated with disfluencies and vocabulary and accentedness is highly correlated with nativeness and pronunciation. Because of the high correlation among rating variables, PCA was performed for dimension reduction. The PCA indicated that all rating factors contributed similarly to the PC1, which was referred to the general impression of oral proficiency. Rating

variables were grouped into two categories to indicate fluency and accent by the PC2. One of the groups consisting of fluency, disfluency, vocabulary, grammar, and comprehensibility variables is more relevant to the knowledge of the language, which affects the formation of the intended message as presented in the Levelt model. This suggests that deficient linguistic knowledge may affect speech planning and the degree of oral fluency. The other rating variables (nativeness, accentedness, and pronunciation) belong to the other group and represent the sound-related factors correlated with the perception of foreign accents. From the weighting contributed by the variables to the PC2 and PC3, the PC2 represented the color of nativeness and accentedness in speech, while the PC3 stood for the characteristics of oral fluency. The multiple regression analysis further indicated that disfluencies and vocabulary have a considerable causal effect on fluency and nativeness has more predictive power on accentedness than on pronunciation.

Accent is the perceptual distance between the speech of listeners and that of speakers. The native listeners from Taiwan recognized a dialectal accent in the speech produced by native speakers from mainland China. Hence, raters ranked the dialectal accent lower because of the differences from their own speech. If the raters were from China, we would expect the opposite scoring patterns. We expected native speakers to receive high scores on all questions; however, their accent scores were low. Several native speakers received low scores on accent and native speakers were from mainland China, whereas the raters were from Taiwan. For accents in L2 speech, the results indicated that improving pronunciation is easier, whereas changing impressions of accents is difficult. Accent scores were generally lower than pronunciation scores.

This study advances the understanding of L2 fluency and foreign-accented speech using the perceptual perspective. This study also has implications for language teaching for designing a corrective method to reduce accents and improve fluency.

References

Anderson-Hsieh, Janet, Ruth Johnson and Kenneth Koehler. 1992. The relationship between native speaker judgments of nonnative pronunciation and deviance in segmentals, prosody, and syllable structure. *Language Learning* 42.4: 529-555.

de Bot, Kees. 1992. A bilingual production model: Levlet's speaking' model adapted. *Applied linguistics* 13.1: 1-24.

Levelt, Willem. J. M. 1989. *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.

Lennon, Paul. 1990. Investigating fluency in EFL: A quantitative approach. *Language Learning* 40.3: 387-417.

Lennon, Paul. 2000. The lexical element on spoken second language fluency. *Perspectives on Fluency*, ed. by Heidi Riggenbach, 25-42. Ann Arbor: University of Michigan Press.

Hosoda, Megumi, and Eugene Stone-Romero. 2010. The effects of foreign accents on employment-related decisions. *Journal of Managerial Psychology* 25.2: 113-132.

Hellwig, Birgit. (n.d.). Elan-linguistic annotator. http://www.lat-mpi.eu/tools/elan/.

Lindemann, Stephanie. 2005. Who speaks "broken English"? US undergraduates' perceptions of non-native English. *International Journal of Applied Linguistics* 15.2: 187-212.

Lev-Ari, Shiri, and Baz Keysar. 2010. Why don't we believe non-native speakers? The influence of accent on credibility. *Journal of Experimental Social Psychology* 46: 1093-1096.

Munro, Murray J. 1998. The effects of noise on the intelligibility of foreign-accented speech. *Studies in Second Language Acquisition* 20.2: 139-154.

Munro, Murray J., and Tracey M. Derwing. 2001. Modeling perceptions of the accentedness and comprehensibility of L2 speech. *Studies in Second Language Acquisition* 23: 451-468.

Munro, Murray. J., and Tracey M. Derwing. 1998. The effects of speaking rate on listener evaluations of native and foreign-accented speech. *Language Learning* 48.2: 159-182.

Munro, Murray. J., Tracey M. Derwing and Susan L. Morton. 2006. The mutual intelligibility of L2 speech. *Studies in Second Language Acquisition* 28.1: 111-131.

Trofimovich, Pavel, and Wendy Baker. 2006. Learning second language suprasegmentals: Effect of L2 experience on prosody and fluency characteristics of L2 speech. *Studies in Second Language Acquisition* 28: 1-30.

Shih, Chilin. 2006. The language class as a community: A task design for speaking proficiency training. *Journal of Chinese Language Teachers Association* 41.2: 1-22.

Shih, Chilin, and Chen-huei Wu. 2011. Evaluating second language fluency. *Proceedings of New Tools and Methods for Very-Large-Scale Phonetic Research*. Philadelphia, USA.

Wu, Chen-huei. 2011. The evaluation of second language fluency and foreign accent. Unpublished doctoral dissertation, University of Illinois at Urbana-Champaign.

[審查：2012.9.4　修改：2013.3.20　接受：2013.5.8]

吳貞慧
Chen-huei Wu
300 國立新竹教育大學 中國語文學系
Dept. of Chinese Language and Literature,
National Hsinchu University of Education
No. 521, Nanda Rd, Hsinchu City, Taiwan 30014
chenhueiwu@gmail.com

從聽覺感知探討第二語言口語流暢度及腔調

吳貞慧
國立新竹教育大學中國語文學系

摘要

本研究採用聽覺評分方式探討華語為第二語言之口語流暢度及腔調的現象。語料為隨機選取自華語學習者自然語料庫的課堂語料，評分項目包括口語流暢度、母語化程度、腔調、不流暢現象、發音、文法正確性、字彙選擇及理解度等。相關性分析顯示口語流暢度與不流暢現象及字彙選擇最為相關；腔調則與發音最為相關。主成分分析顯示第一主成分代表口語表達的整體印象分數；第二主成份將變項分成兩組：口語流暢度與形成語言的知識要素相關；腔調與語音相關，而第二主成分主要是由母語化程度及腔調組成；第三主成分則代表口語流暢度。複迴歸分析進一步驗證不流暢現象與字彙對於預測口語流暢度有顯著的影響，而母語化程度則是造成腔調的主要變項。

關鍵字：第二語言習得　口語流暢度　腔調　口音　聽覺實驗　華語